

2-409
ONLINE SEARCH REQUEST FORM

USER B. CELS4SERIAL NUMBER 07/822043ART UNIT 1811PHONE 308-4001DATE 02/24/93

Please give a detailed statement of requirements. Describe as specifically as possible the subject matter to be searched. Define any terms that may have special meaning. Give examples or relevant citations, authors, or keywords, if known.

You may include a copy of the broadest and or relevant claim(s).

PROTEIN WHICH IS A GROWTH OR MOTILITY FACTOR WAS ISOLATED FROM A2055 HUMAN MELANOMA CELLS. THREE PROTEINS WERE CALLED AUTOTAXIN AND HAVE MW OF 120-130 kDa. THE PROTEINS WERE ENZYMATICALLY DEGRADED FORMING PEPTIDES OF SEQ ID NO 1-11 AND 20-33.

PLEASE SEARCH THE FOLLOWING:

I CAS EXACT SEQUENCE SEARCH FOR SEQ ID NO¹ 1-11 AND 20-33.

- LIST References dated Before 1992

II CAS EXPAND SEQUENCE SEARCH FOR SEQ ID NO¹ 1-11 AND 20-33

- LIST ¹References dated Before 1992.

III NARROW ALL SEARCH QUERIES OF II IN CAS USING FOLLOWING WORD SEARCHES:

A. (120 OR 121 OR 122 OR 123 OR 124 OR 125 OR 126 OR 127 OR 128 OR 129 OR 130)
(W) (100 OR 110 OR 120 OR 130 OR 140 OR 150 OR 160 OR 170 OR 180 OR 190) / BI, AB

B. INVENTOR SEARCH

C. (MOTILITY) (W) (FACTOR#) / BI, AB

D. (GROWTH) (W) (FACTOR#) / BI, AB

E. (TUMOR?) / BI, AB

F. A2055

- LIST References DATED PRIOR TO 1992.

INCLUDE ABSTRACT + LISTING OF PEPTIDES WITH SEQ ID NO¹ 1-11 AND 20-33

STAFF USE ONLY

COMPLETED _____

SEARCHER _____

ONLINE TIME _____ TOTAL TIME _____
(in minutes)

NO. OF DATABASES _____

SYSTEMS

____ CAS ONLINE
____ DARC/QUESTEL
____ DIALOG
____ SDC
____ OTHER

643
48-10

Celsa 822043

=> d his

(FILE 'REGISTRY' ENTERED AT 09:02:07 ON 25 FEB 93)
DEL HIS

FILE 'CA' ENTERED AT 09:02:40 ON 25 FEB 93

L1 0 S AUTOTAXIN#
L2 0 S L1/AB
L3 48 S A2058/AB, BI
L4 7062 S MELANOMA/AB, BI
L5 39 S L4 AND L3
E STRACKE, M/AU
L6 12 S E3 OR E6
L7 12 S LIOTTA, L/AU
E LIOTTA, L/AU
L8 163 S E5-9 OR L7
E SCHIFFMANN, E/AU
L9 67 S E3-6
L10 222 S L6 OR L8 OR L9 *all three authors*
L11 22 S L10 AND L5

FILE 'REGISTRY' ENTERED AT 09:07:02 ON 25 FEB 93

L12 87 S VLNYF|YLNAT|YPAFK|QAEVS|WHVAAN/SQSP
L13 0 S YDVPWNETI|SPPFENINLY|GGQPLWITATK|QYLHQYGGSS|HLLYGRPAVLY|
L14 0 S VNMQTVFVGYGPTGFK|DIEHLSLDFFR|TEFLSNYLTNVDDITLVPGTGR|Y
L15 0 S PEEVT:PNYL|P:LDVYK/SQSP
L16 ~~(0 S PEEVT:PNYL|P:LDVYK/SQSP~~
L17 0 S L12 AND SQL=<6

FILE 'CA' ENTERED AT 09:15:40 ON 25 FEB 93

L18 60 S L12 OR L12/D
L19 60 S (MOTILITY FACTOR#)/AB, BI
L20 31402 S (GROWTH FACTOR#)/AB, BI
L21 103596 S (TUMOR# OR TUMOUR#)/AB, BI
L22 0 S L18 AND (L19 OR L20 OR L21 OR L3)
~~L23 0 S L22 AND L10~~
~~L24 108 S L21 AND L10~~
L25 0 S L18 AND L10
L26 0 S L18 AND L4
L27 58706 S (KD OR KDA OR K DALTON# OR KILODALTON#)/AB, BI
L28 19 S L18 AND L27
L29 330070 S (121 OR 120 OR 122 OR 123 OR 124 OR 125 OR 126 OR 127 O
L30 3872 S L27 (L) L29
L31 1 S L30 AND L18
L32 0 S L18 AND (L29) (W) 000/AB, BI
L33 58 S L18 NOT 1992/PY

*This includes all
the peptides*

FILE 'REGISTRY' ENTERED AT 09:24:44 ON 25 FEB 93
L34 2 S L12 AND SQL=<100

FILE 'CA' ENTERED AT 09:25:23 ON 25 FEB 93

Celsa 822043

L35 2 S L34 OR L34/D
L36 12 S L11 AND (PROTEIN# OR PEPTIDE# OR POLYPEPTIDE#)

=> d bib ab hitrn l31

L31 ANSWER 1 OF 1 COPYRIGHT 1993 ACS

AN CA109(19):164814p

TI Isolation and structural characterization of the human 4F2 heavy-chain gene, an inducible gene involved in T-lymphocyte activation

AU Gottesdiener, Keith M.; Karpinski, Beverly A.; Lindsten, Tullia; Strominger, Jack L.; Jones, Nancy H.; Thompson, Craig B.; Leiden, Jeffrey M.

CS Howard Hughes Med. Inst., Univ. Michigan

LO Ann Arbor, MI 48109, USA

SO Mol. Cell. Biol., 8(9), 3809-19

SC 3-3 (Biochemical Genetics)

SX 13, 15

DT J

CO MCEBD4

IS 0270-7306

PY 1988

LA Eng

AB The human 4F2 cell surface antigen is a 120-

kilodalton (kDa) disulfide-linked heterodimer

which is composed of an 80- to 90-kDa glycosylated heavy chain (4F2HC) and a 35- to 40-kDa nonglycosylated light chain (4F2LC). 4F2 belongs to a family of inducible cell surface mols. which are involved in T-lymphocyte activation and growth. To better understand the mol. mechanism(s) that controls 4F2HC gene expression in both resting and activated T cells, a 4F2HC human genomic clone was isolated and structurally characterized. The 4F2HC gene spans 8 kilobases of chromosome 11 and is composed of nine exons. The 5' upstream region of the gene displays several properties which are characteristic of housekeeping genes. It is G + C rich and hypomethylated in peripheral blood lymphocyte DNA and contains multiple binding sites for the Sp1 transcription factor while lacking TATA or CCAAT sequences. This region of the gene also displays sequence homologies with several other inducible T-cell genes, including the interleukin-2, interleukin-2 receptor .alpha. chain, dihydrofolate reductase, thymidine kinase, and transferrin receptor genes. A 255-base-pair fragment of the 4F2HC gene which contains 154 base pairs of the 5' flanking sequence was able to efficiently promote expression of the bacterial chloramphenicol acetyltransferase gene in human Jurkat T cells, indicating that it contains promoter or enhancer (or both) sequences. Analyses of chromatin structure in resting and lectin-activated T cells revealed the presence of stable DNase I-hypersensitive sites within both the 5' flanking and intron 1 regions of the 4F2HC gene. Although the 4F2HC gene displayed many of the structural features characteristic of a constitutively expressed gene, lectin-mediated activation of

QH 506 Mb

Celsa 822043

resting peripheral blood T lymphocytes resulted in a dramatic increase in steady-state levels of 4F2HC mRNA.

IT 111309-64-9

(amino acid sequence of)

=> d bib abs hitrn 135 1-2

L35 ANSWER 1 OF 2 COPYRIGHT 1993 ACS

AN CA114(7):57704j

TI The primary structure of DNA binding protein II from the extreme thermophilic bacterium *Thermus thermophilus*

AU Zierer, Rainer; Choli, Dora

CS Abt. H-G Wittmann, Max Planck Inst. Mol. Genet.

LO Berlin, Fed. Rep. Ger.

SO FEBS Lett., 273(1-2), 59-62

SC 6-3 (General Biochemistry)

DT J

CO FEBLAL

IS 0014-5793

PY 1990

LA Eng

AN CA114(7):57704j

AB The primary structure of DNA binding protein II (DNA bp II) from the extreme thermophilic bacterium *T. thermophilus* has been established by a combination of manual and automated techniques. The protein has 95 residues and a mol. mass of 11,843. Comparison of the primary structure with the known sequence data of DNA bp II from *Clostridium pasteurianum*, *Bacillus stearothermophilus*, *Escherichia coli*, *Rhizobium meliloti*, *Anabena*, *Thermoplasma acidophilum*, *Pseudomonas aeruginosa*, and *Bacillus caldolyticus* reveals a clear homol. among these small basic proteins. In particular 2 short sequences in the middle and C-terminal part of the proteins (positions 46-51 and 63-65, resp.) are completely conserved.

IT 131571-34-1, Protein II (*Thermus thermophilus* strain HB8

DNA-binding)

(amino acid sequence of)

L35 ANSWER 2 OF 2 COPYRIGHT 1993 ACS

AN CA111(18):160184b

TI Outer membrane proteins and nucleotide sequences coding for them in vaccines and diagnostic assays for *Haemophilus influenzae*

AU Deich, Robert A.; Zlotnick, Gary; Green, Bruce

CS Praxis Biologics, Inc.

LO USA

SO PCT Int. Appl., 129 pp.

PI WO 8804932 A1 14 Jul 1988

DS W: AU, DK, JP, KR

RW: AT, BE, CH, DE, FR, GB, IT, LU, NL, SE

AI WO 87-US3423 23 Dec 1987

PRAI US 86-948364 31 Dec 1986

US 87-20849 2 Mar 1987

US 87-132073 11 Dec 1987
 IC ICM A61K035-66
 ICS A61K039-00; G01N033-53; G01N033-569
 SC 63-3 (Pharmaceuticals)
 SX 3, 15
 DT P
 CO PIXXD2
 PY 1988
 LA Eng
 AN CA111(18):160184b
 AB

Peptides and proteins related to an epitope comprising an outer membrane protein (OMP) of *H. influenzae* are described. The peptides and proteins can be prepd. by methods including novel and improved methods of purifn. from *H. influenzae* cultures, and by recombinant DNA and chem. synthetic techniques. Addnl., recombinant vectors contg. nucleotide sequences encoding PBOMP-1 (Praxis Biologics outer membrane protein) and PBOMP-2 related peptides and proteins are also described. Recombinant vectors include plasmid DNA and viral DNA such as human viruses, animal viruses, insect viruses and bacteriophages that direct the expression of the PBOMP-1 and PBOMP-2 related peptides and proteins in appropriate host cells. The peptides, proteins and viruses both live and inactivated are used as immunogens in vaccine formulations to protect against *H. influenzae* infections. The peptides and proteins are also used as reagents in immunoassays as well as to prep. Igs for passive immunization. Use of the nucleotide sequences encoding the PBOMP related peptides and proteins in hybridization assays is also described. *H. influenzae* Eagan was grown overnight in brain heart infusion medium contg. 10 .mu.g hemin/mL and centrifuged, and the pellet was suspended in 10 mM HEPES-NaOH (pH 7.4) 1 mM EDTA and sonicated (2 repetitions). After addn. of NaCl to 0.5 M and ultracentrifugation at 100,000 .times. g for 1 h, the membrane pellet was repeatedly extd. with 1% sarcosyl in the same buffer as above and then centrifuged, suspended in 50 mM Tris (pH 8.0), 5 mM EDTA and extd. with octylglucoside (2.times.) and sarcosyl (2.times.) (both 1% in 50 mM Tris, 5 mM EDTA, pH 8.0). After centrifugation at 100,000 .times. g, the pellet was solubilized by heating at 60.degree. for 1 h in the presence of detergent (e.g. Tween 80, CHAPS) and then subjected to SDS-PAGE. PBOMP-1 bands were excised from the gel and minced, and the gel fragments were injected (20 .mu.g PBOMP-1) into New Zealand white rabbits at 0, 2, and 3 wks, with bimonthly boost with the same amt. Infant Sprague-Dawley rats (4 days old) were inoculated with 0.1 mL of varying dilns. of the rabbit antisera, and at 18 h postimmunization were challenged with 104-106 cells of *H. influenzae* type b strains HST-60, HST-61, and Eagan. At 72 h postchallenge, a 1/30 diln. of the antisera protected 100% of the rats.

IT 122983-86-2

(peptide, of outer membrane protein of *Haemophilus influenzae*)

=> d bib 136 1-12

- L36 ANSWER 1 OF 12 COPYRIGHT 1993 ACS
 AN CA114(25):243494r
 TI A novel matrix metalloproteinase inhibitor and a cDNA clone for it
 AU Stetler-Stevenson, William G.; Liotta, Lance A.; Krutzsh, Henry
 CS United States Dept. of Commerce
 LO USA
 SO PCT Int. Appl., 47 pp.
 PI WO 9011287 A1 4 Oct 1990
 DS W: AU, CA, JP
 RW: AT, BE, CH, DE, DK, ES, FR, GB, IT, LU, NL, SE
 AI WO 90-US1526 21 Mar 1990
 PRAI US 89-326334 21 Mar 1989
 US 89-380431 17 Jul 1989
 US 89-395453 18 Aug 1989
 IC C07H015-12; C12N001-22; C12N015-15; A61K037-02; A61K039-00;
 A01W037-18
 SC 7-3 (Enzymes)
 SX 3
 DT P
 CO PIXXD2
 PY 1990
 LA Eng
- L36 ANSWER 2 OF 12 COPYRIGHT 1993 ACS
 AN CA114(23):221378d
 TI Type IV (pro)collagenase-derived **peptides** as
 metalloproteinase inhibitors, antibodies to such **peptides**,
 and use of the **peptides** and antibodies in the treatment
 and diagnosis of cancer and other diseases
 AU Liotta, Lance A.; Stetler-Stevenson, William; Krutzsch, Henry
 CS National Institutes of Health
 LO USA
 SO U. S. Pat. Appl., 44 pp. Avail. NTIS Order No. PAT-APPL-6-317 407.
 PI US 317407 A0 15 Jul 1990
 AI US 89-317407 1 Mar 1989
 SC 1-6 (Pharmacology)
 SX 7, 9
 DT P
 CO XAXXAV
 PY 1990
 LA Eng
- L36 ANSWER 3 OF 12 COPYRIGHT 1993 ACS
 AN CA114(5):40157k
 TI Tissue inhibitor of metalloproteinases-2 (TIMP-2) mRNA expression in
 tumor cell lines and human tumor tissues
 AU Stetler-Stevenson, William G.; Brown, Peter D.; Onisto, Maurizio;
 Levy, Anna T.; Liotta, Lance A.
 CS Lab. Pathol., Natl. Cancer Inst.
 LO Bethesda, MD 20892, USA
 SO J. Biol. Chem., 265(23), 13933-8

SC 14-1 (Mammalian Pathological Biochemistry)
 SX 3
 DT J
 CO JBCHA3
 IS 0021-9258
 PY 1990
 LA Eng

L36 ANSWER 4 OF 12 COPYRIGHT 1993 ACS
 AN CA114(3):18521c
 TI Cloning and characterization of human tumor cell interstitial
 collagenase
 AU Templeton, Nancy Smyth; Brown, Peter D.; Levy, Anna T.; Margulies,
 Inger M. K.; Liotta, Lance A.; Stetler-Stevenson, William G.
 CS Lab. Pathol., Natl. Cancer Inst.
 LO Bethesda, MD 20892, USA
 SO Cancer Res., 50(17), 5431-7
 SC 3-3 (Biochemical Genetics)
 SX 7, 14
 DT J
 CO CNREA8
 IS 0008-5472
 PY 1990
 LA Eng

L36 ANSWER 5 OF 12 COPYRIGHT 1993 ACS
 AN CA113(13):111183h
 TI Sulfatide-binding domain of the laminin A chain
 AU Taraboletti, Giulia; Rao, C. N.; Kruttsch, Henry C.; Liotta, Lance
 A.; Roberts, David D.
 CS Lab. Pathol., Natl. Cancer Inst.
 LO Bethesda, MD 20892, USA
 SO J. Biol. Chem., 265(21), 12253-8
 SC 6-3 (General Biochemistry)
 SX 13
 DT J
 CO JBCHA3
 IS 0021-9258
 PY 1990
 LA Eng

L36 ANSWER 6 OF 12 COPYRIGHT 1993 ACS
 AN CA112(13):112781t
 TI Autocrine motility factor stimulates a three-fold increase in
 inositol trisphosphate in human melanoma cells
 AU Kohn, Elise C.; Liotta, Lance A.; Schiffmann, Elliott
 CS Med. Branch, Natl. Cancer Inst.
 LO Bethesda, MD 20892, USA
 SO Biochem. Biophys. Res. Commun., 166(2), 757-64
 SC 2-10 (Mammalian Hormones)
 DT J

CO BBRCA9
IS 0006-291X
PY 1990
LA Eng

L36 ANSWER 7 OF 12 COPYRIGHT 1993 ACS
AN CA112(3):17896a
TI The elastin receptor shows structural and functional similarities to
the 67-kDa tumor cell laminin receptor
AU Mecham, Robert P.; Hinek, Aleksander; Griffin, Gail L.; Senior,
Robert M.; Liotta, Lance A.
CS Dep. Med., Jew. Hosp.
LO St. Louis, MO 63110, USA
SO J. Biol. Chem., 264(28), 16652-7
SC 6-3 (General Biochemistry)
DT J
CO JBCHA3
IS 0021-9258
PY 1989
LA Eng

L36 ANSWER 8 OF 12 COPYRIGHT 1993 ACS
AN CA111(25):225947z
TI The type I insulin-like growth factor is a motility receptor in
human melanoma cells
AU Stracke, Mary L.; Engel, Jason D.; Wilson, Lori W.; Rechler, Matthew
M.; Liotta, Lance A.; Schiffman, Elliott
CS Lab. Pathol., Natl. Cancer Inst.
LO Bethesda, MD 20892, USA
SO J. Biol. Chem., 264(36), 21544-9
SC 2-10 (Mammalian Hormones)
SX 14, 15
DT J
CO JBCHA3
IS 0021-9258
PY 1989
LA Eng

L36 ANSWER 9 OF 12 COPYRIGHT 1993 ACS
AN CA110(17):150420r
TI The activation of human type IV collagenase proenzyme. Sequence
identification of the major conversion product following
organomercurial activation
AU Stetler-Stevenson, William G.; Kruttsch, Henry C.; Wachter, Mary P.;
Margulies, Inger M. K.; Liotta, Lance A.
CS Lab. Pathol., Natl. Cancer Inst.
LO Bethesda, MD 20892, USA
SO J. Biol. Chem., 264(3), 1353-6
SC 7-5 (Enzymes)
DT J
CO JBCHA3

Celsa 822043

IS 0021-9258
PY 1989
LA Eng

L36 ANSWER 10 OF 12 COPYRIGHT 1993 ACS
AN CA108(11):92970r
TI Autocrine motility factor (AMF) formation by cancer cells, its
determination in cancer diagnosis, and AMF inhibitors for cancer
treatment
AU Liotta, L. A.; Schiffmann, E.
CS United States Dept. of Health and Human Services
LO USA
SO U. S. Pat. Appl., 30 pp. Avail. NTIS Order No. PAT-APPL-7-58381.
PI US 58381 A0 1 Nov 1987
AI US 87-58381 5 Jun 1987
SC 15-5 (Immunochemistry)
DT P
CO XAXXAV
PY 1987
LA Eng

L36 ANSWER 11 OF 12 COPYRIGHT 1993 ACS
AN CA107(13):110769t
TI Pertussis toxin inhibits stimulated motility independently of the
adenylate cyclase pathway in human melanoma cells
AU Stracke, Mary L.; Guirguis, Raouf; Liotta, Lance A.; Schiffmann,
Elliott
CS Lab. Pathol., Natl. Inst. Health
LO Bethesda, MD 20892, USA
SO Biochem. Biophys. Res. Commun., 146(1), 339-45
SC 4-5 (Toxicology)
DT J
CO BBRCA9
IS 0006-291X
PY 1987
LA Eng

L36 ANSWER 12 OF 12 COPYRIGHT 1993 ACS
AN CA105(7):58748c
TI Tumor cell autocrine motility factor
AU Liotta, Lance A.; Mandler, Raya; Murano, Genesio; Katz, David A.;
Gordon, Richard K.; Chiang, Peter K.; Schiffmann, Elliott
CS Natl. Cancer Inst., Food Drug Adm.
LO Bethesda, MD 20892, USA
SO Proc. Natl. Acad. Sci. U. S. A., 83(10), 3302-6
SC 14-1 (Mammalian Pathological Biochemistry)
DT J
CO PNASA6
IS 0027-8424
PY 1986
LA Eng

=> d bib ab hitrn l33 1-10

L33 ANSWER 1 OF 58 COPYRIGHT 1993 ACS
 AN CA118(1):2872q
 TI Cloning, sequencing and expression of the sialidase gene from
 Actinomyces viscosus DSM 43798
 AU Henningsen, Michaela; Roggentin, Peter; Schauer, Roland
 CS Biochem. Inst., Univ. Kiel
 LO Kiel W-2300, Germany
 SO Biol. Chem. Hoppe-Seyler, 372(12), 1065-72
 SC 7-5 (Enzymes)
 SX 3, 10
 DT J
 CO BCHSEI
 IS 0177-3593
 PY 1991
 LA Eng
 AB Chromosomal DNA from *A. viscosus* was digested with restriction
 endonucleases and the fragments ligated with pUC-vectors were used
 to transform *Escherichia coli* cells. Clones bearing the required
 sialidase gene were detected by spraying the colonies with the
 fluorogenic sialidase substrate 4-methylumbelliferyl-.alpha.-D-
 acetylneuraminic acid. The identity of the cloned sialidase was
 confirmed after 5700-fold enrichment and comparison with the
 purified enzyme of *A. viscosus*. Both sialidases were identical with
 regard to mol. mass, substrate specificity tested with
 sialyllactoses, and the inhibition of their activity by heterologous
 antisialidase antibodies. The sequenced insert revealed a mol% G + C
 of 68.2, typical for *A. viscosus*. An open reading frame of 2739 bp
 follows a sequence with dyad symmetry and an AG-rich region, and
 codes for 913 amino acids representing a mol. mass of 113 kDa. The
 conserved amino acid sequence [Ser-X-Asp-X-Gly-X-Thr-Trp] typical
 for bacterial sialidases was found at 5 positions in the predicted
 amino acid sequence. The gene of this enzyme is expressed by *E.*
coli, despite the low relatedness of the 2 species.
 IT 144813-79-6, Sialidase (*Actinomyces viscosus* DSM 43798
 reduced)
 (amino acid sequence of, complete)

L33 ANSWER 2 OF 58 COPYRIGHT 1993 ACS
 AN CA117(17):167649q
 TI Peptide composition and enzyme activities of isolated pyrenoids from
 the green alga *Bryopsis maxima*
 AU Okada, M.; Okabe, Y.; Kono, M.; Nakayama, K.; Satoh, H.
 CS Fac. Sci., Toho Univ.
 LO Funabashi 274, Japan
 SO Can. J. Bot., 69(5), 1053-61
 SC 11-1 (Plant Biochemistry)
 SX 3, 7
 DT J

CO CJBOAW
 IS 0008-4026
 PY 1991
 LA Eng
 AB Pyrenoids of *B. maxima* contained several minor components other than the large subunit (LS) and the small subunit of ribulose 1,5-bisphosphate carboxylase/oxygenase (Rubisco). Among the minor components, polypeptides of 95, 67, and 41 kDa reacted with an antibody against the LS polypeptide. Amino acid sequences of these polypeptides were detd. and compared with that deduced from the LS gene (*rbcl*) screened from the chloroplast DNA library of *B. maxima*. The N-terminal sequence of the LS peptide was not post-translationally processed and was almost identical with those of the polypeptides of 91, 67, and 41 kDa. The starch grains surrounding the pyrenoids contained a polypeptide of 66 kDa that was assigned as starch synthase.

IT 143861-32-9 143861-35-2
 (amino acid sequence of)

L33 ANSWER 3 OF 58 COPYRIGHT 1993 ACS
 AN CA117(15):144555s
 TI Nucleotide sequence of the large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase from the green alga *Bryopsis maxima*
 AU Kono, Midori; Satoh, Hiroyuki; Okabe, Yasuyuki; Abe, Yasuko; Nakayama, Katsumi; Okada, Mitsumasa
 CS Fac. Sci., Toho Univ.
 LO Funabashi 274, Japan
 SO Plant Mol. Biol., 17(3), 505-8
 SC 3-3 (Biochemical Genetics)
 SX 7, 10, 11
 DT J
 CO PMBIDB
 IS 0167-4412
 PY 1991
 LA Eng
 AB A library was constructed for whole chloroplast DNA of *B. maxima*. The library, screened with the DNA probe of the large subunit (LS) of cyanobacterium Rubisco, showed pos. hybridization to 50 out of approx. 9000 plaques. λ .MHB7, contained the whole *rbcl* gene of *B. maxima*. The nucleotide sequence contained 1425 bp, encoding 475 amino acids. The deduced amino acid sequence of *B. maxima* LS protein showed 92% homol. with that of *Chlorella* LS protein and 90% homol. with that of *Chlamydomonas* LS protein. The amino acid sequences of RuDP-binding sites and CO₂-binding site of *B. maxima* LS protein showed as much of 95% homol. to those of *Chlamydomonas*. An intron of 2467 bp was found between codons 268 and 269 of *B. maxima* *rbcl*. The boundary sequences of the intron most resembled the group III intron described for chloroplast DNA. There was an open reading frame of 822 bp in this intron. It showed no significant gene homol. with the sequences in the GenBank database, but its transcribed products were

detected by Northern blotting using the whole of the reading frame as a probe. Efforts are underway to assess the physiol. significance of this region. A sequence located 118 bp upstream from the B. maxima rbcL initiation site for translation was identical with that of the -10 box of the LS promoter region of rbcL of tobacco, maize and spinach. The sequence T-T-G-G-T-T (at the position of 19 bp upstream from the -10 box) apparently corresponded to the -35 box. A long palindrome sequence (position 4301 to 4346) capable of forming a 23 bp stem was in the rbcL termination region. The partial coding region was found on the same strand at 310 bp upstream from the rbcL coding sequence. Nucleotide sequence homol. was noted between this coding region and that of the .alpha. subunit of ATP synthase (atpA) of liverwort, pea, spinach and tobacco.

IT 143637-87-0 143637-89-2
(amino acid sequence of)

L33 ANSWER 4 OF 58 COPYRIGHT 1993 ACS

AN CA117(11):107156s

TI M13 cloning and nucleotide sequence of the sorghum chloroplast gene for the large subunit of ribulose-1,5-bisphosphate carboxylase

AU Lou, Huan; Zhang, Bin; Qi, Defang

CS Shanghai Inst. Biochem., Acad. Sin.

LO Shanghai, Peop. Rep. China

SO Shengwu Huaxue Yu Shengwu Wuli Xuebao, 21(3), 179-88

SC 7-5 (Enzymes)

SX 3, 11

DT J

CO SHWPAU

IS 0582-9879

PY 1989

LA Ch

AB The cloned gene for the large subunit of ribulose-1,5-bisphosphate carboxylase (rboL) from sorghum has been sequenced by the method of M13/dideoxy chain-termination. Among the total no. of 1839 bp sequenced, the coding region of the rbcL gene contains 1428 bp (476 codons). A sequence GGAGG corresponding to SD sequence occurs 6 to 10 bp upstream from the initiation codon ATG. The 5' flanking region contains sequences resembling the "-10 region" and "-35 region" sequences of the prokaryote promoter. The 3' flanking region contains a sequence which can form a stem-and-loop structure similar to the terminator of the prokaryote gene. Comparison of rboL gene of sorghum with that of maize shows 97.7% homol. in the coding region, and 90% homol. in the noncoding flanking region. However, the rboL genes of sorghum and maize differ obviously from those of C3 plants (spinach and tobacco). In M13 cloning, restriction enzyme partial digestion of clones resulted in identification of recombinant phages more simple, rapid and accurate. The sequencing of that DNA region was nonrandom and the efficiency was improved.

IT 143108-32-1
(amino acid sequence of, complete)

L33 ANSWER 5 OF 58 COPYRIGHT 1993 ACS

AN CA117(9):84434q

TI Expression of M-cadherin, a member of the cadherin multigene family, correlates with differentiation of skeletal muscle cells

AU Donalies, Michael; Cramer, Matthias; Ringwald, Martin; Starzinski-Powitz, Anna

CS Inst. Genet., Univ. Koeln

LO Cologne D-5000/1, Germany

SO Proc. Natl. Acad. Sci. U. S. A., 88(18), 8024-8

SC 3-3 (Biochemical Genetics)

SX 6, 13

DT J

CO PNASA6

IS 0027-8424

PY 1991

LA Eng

AB Cadherins, a multigene family of transmembrane glycoproteins, mediate Ca²⁺-dependent intercellular adhesion. They are thought to be essential for the control of morphogenetic processes, including myogenesis. The identification and characterization of the cDNA of another member of the cadherin family, M-cadherin (M for muscle), from differentiating muscle cells is reported. The longest open reading frame of the cDNAs isolated contains almost the entire coding region of the mature M-cadherin as detd. by sequence homol. to the known cadherins. M-cadherin mRNA is present at low levels in myoblasts and is upregulated in myotube-forming cells. In mouse L cells (fibroblasts), M-cadherin mRNA is undetectable. This expression pattern indicates that M-cadherin is part of the myogenic program and may provide a trigger for terminal muscle differentiation.

IT 142845-03-2, Cadherin M (mouse C-terminal fragment protein moiety reduced)
(amino acid sequence of)

L33 ANSWER 6 OF 58 COPYRIGHT 1993 ACS

AN CA117(7):64022a

TI Molecular cloning and analysis of small optic lobes, a structural brain gene of Drosophila melanogaster

AU Delaney, S. J.; Hayward, D. C.; Barleben, F.; Fischbach, K. F.; Miklos, G. L. Gabor

CS Res. Sch. Biol. Sci., Aust. Natl. Univ.

LO Canberra 2601, Australia

SO Proc. Natl. Acad. Sci. U. S. A., 88(16), 7214-18

SC 3-3 (Biochemical Genetics)

SX 6, 12

DT J

CO PNASA6

IS 0027-8424

PY 1991

LA Eng

AB Mutations in the small optic lobes (sol) gene of D. melanogaster

cause specific cells to degenerate in the developing optic lobes, resulting in the absence of certain classes of columnar neurons. These neuronal defects lead to specific alterations in behavioral characteristics, particularly during flight and walking maneuvers. The wild-type sol locus was isolated by microcloning and chromosomal walking and its genetic and mol. limits were established. Two major transcripts of 5.8 and 5.2 kb are produced from this locus by alternative splicing and are present throughout the entire life cycle. Sequence analyses of cDNAs corresponding to these 2 classes of transcripts predict 2 proteins of 1597 and 395 amino acids. The first shows similarity in its carboxyl-terminal region to the catalytic domain of a vertebrate calcium-activated neutral protease (calpain), whereas its amino-terminal region contains several zinc-finger-like repeats of the form WXC₂Y₂CX₁₀₋₁₁CX₂C. The second predicted protein contains only the first 2 of the zinc-finger-like repeats and is missing the calpain domain. By constructing transgenic flies carrying a single wild-type copy of the sol gene in a homozygous sol mutant background, the normal neuroanatomical phenotype was restored to individuals that would have developed mutant brains.

IT 142661-94-7, Protein (Drosophila melanogaster clone
 .lambda.c0.22/.lambda.c0.32 gene small-optic-lobes reduced)
 (amino acid sequence of)

L33 ANSWER 7 OF 58 COPYRIGHT 1993 ACS

AN CA116(13):122076p

TI Tick-borne encephalitis virus strain 205: nucleotide sequence of
 genes and complete amino acid sequence of viral proteins

AU Safronov, P. F.; Netesov, S. V.; Mikryukova, T. P.; Blinov, V. M.;
 Osipova, E. G.; Kiseleva, N. N.; Sandakhchiev, L. S.

CS VNII Mol. Biol.

LO USSR

SO Mol. Genet., Mikrobiol. Virusol., (4), 23-9

SC 3-2 (Biochemical Genetics)

SX 6

DT J

CO MGMVDU

IS 0208-0613

PY 1991

LA Russ

AB The 10466 nucleotide long sequence of the cDNA copy of the
 tick-borne encephalitis strain 205 viral genome was detd. It
 includes the 5'-nontranslated region, the genes for structural as
 well as nonstructural proteins and the first 93 nucleotides of
 3'-nontranslated region. The difference in amino acid sequences of
 structural and nonstructural proteins of strains 205, Sofjin and
 Neudoerfl of the tick-borne encephalitis virus and the nucleotide
 changes in 5'- and 3'-nontranslated regions of these strains are
 discussed.

IT 128808-03-7, Protein M (tick-borne encephalitis virus strain 205
 clone II-71/I-62) 128808-67-3, Glycoprotein E (tick-borne

encephalitis virus strain 205 clone II-71/I-62 protein moiety reduced) 128808-68-4, Glycoprotein prM (tick-borne encephalitis virus strain 205 clone II-71/I-62 precursor protein moiety reduced) 128808-69-5, Glycoprotein prM (tick-borne encephalitis virus strain 205 clone II-71/I-62 protein moiety reduced) 139074-91-2, Protein C (tick-borne encephalitis virus strain 205 clone II-71)
139076-24-7 139076-26-9, Protein NS 1 (tick-borne encephalitis virus strain 205 clone III-23/I-3 reduced) 139076-27-0, Protein NS 5 (tick-borne encephalitis virus strain 205 clone II-4/III-45/III-38/VI-49 reduced) **139076-28-1**, Protein NS 3 (tick-borne encephalitis virus strain 205 clone I-47/III-30/I-30 reduced) 139076-29-2, Protein NS 2A (tick-borne encephalitis virus strain 205 clone I-3 reduced) 139076-30-5, Protein NS 4A (tick-borne encephalitis virus strain 205 clone I-30 reduced) 139076-31-6, Protein NS 2B (tick-borne encephalitis virus strain 205 clone I-3/III-59 reduced) 139076-32-7, Protein NS 4B (tick-borne encephalitis virus strain 205 clone I-30/III-15 reduced) (amino acid sequence of)

L33 ANSWER 8 OF 58 COPYRIGHT 1993 ACS

AN CA116(1):1417s

TI The regions of sequence variation in caulimovirus gene VI

AU Sanger, Margaret; Daubert, Steve; Goodman, Robert M.

CS Dep. Plant Pathol., Univ. California

LO Davis, CA 95616, USA

SO Virology, 182(2), 830-4

SC 3-2 (Biochemical Genetics)

SX 6

DT J

CO VIRLAX

IS 0042-6822

PY 1991

LA Eng

AB The sequence of gene VI from figwort mosaic virus (FMV) clone x4 was detd. and compared with that previously published for FMV clone DxS. Both clones originated from the same virus isolation, but the virus used to clone DxS was propagated extensively in a host of a different family prior to cloning whereas that used to clone x4 was not. Differences in the amino acid sequence inferred from the DNA sequences occurred in 2 clusters. An N-terminal conserved region preceded 2 regions of variation sepd. by a central conserved region. Variation in cauliflower mosaic virus (CaMV) gene VI sequences, all of which were derived from virus isolates from hosts from one host family, was similar to that seen in the FMV comparison, though the extent of variation was less. Alignment of gene VI domains from FMV and CaMV revealed regions of amino acid sequence identical in both viruses within the conserved regions. The similarity in the pattern of conserved and variable domains of these two viruses suggests common host-interactive functions in caulimovirus gene VI homologs, and possibly an analogy between caulimoviruses and certain animal viruses in the influence of the host on sequence variability of

viral genes.

IT 137800-99-8, Protein IBMP (figwort mosaic virus clone DxS reduced) 137801-00-4, Protein IBMP (figwort mosaic virus clone x4 reduced)
(amino acid sequence of)

L33 ANSWER 9 OF 58 COPYRIGHT 1993 ACS

AN CA115(19):199835x

TI Complete cDNA sequence of a South American isolate of potato virus X [Erratum to document cited in CA114(1):1317t]

AU Orman, Betina, E.; Celnik, Rosana M.; Mandel, Alejandra M.; Torres, Hector N.; Mentaberry, Alejandro N.

CS Inst. Invest. Ing. Genet. Biol. Mol., CONICET

LO Buenos Aires 1428, Argent.

SO Virus Res., 19(2-3), 236

SC 3-2 (Biochemical Genetics)

SX 6

DT J

CO VIREDF

IS 0168-1702

PY 1991

LA Eng

AB Errors in the genomic sequence in Figure 2 have been cor. The errors were reflected in the index entries.

IT 123514-40-9, Protein (potato virus X strain Xc coat reduced) 130843-48-0, Protein (potato virus X strain Xc 7.6-kilodalton reduced) 130844-13-2, Protein (potato virus X strain Xc 12.3-kilodalton reduced) 130844-14-3, Protein (potato virus X strain Xc 165-kilodalton reduced) 130844-15-4, Protein (potato virus X strain Xc 24.3-kilodalton reduced)
(amino acid sequence of (Erratum))

L33 ANSWER 10 OF 58 COPYRIGHT 1993 ACS

AN CA115(15):152140b

TI The recognition component of the N-end rule pathway

AU Bartel, Bonnie; Wuenning, Ingrid; Varshavsky, Alexander

CS Dep. Biol., Massachusetts Inst. Technol.

LO Cambridge, MA 02139, USA

SO EMBO J., 9(10), 3179-89

SC 3-3 (Biochemical Genetics)

SX 6

DT J

CO EMJODG

IS 0261-4189

PY 1990

LA Eng

AB The N-end rule-based degrdn. signal, which targets a protein for ubiquitin-dependent proteolysis, comprises a destabilizing amino-terminal residue and a specific internal lysine residue. Here, the isolation and functional anal. of a gene (UBR1) for the N-end recognizing protein of the yeast *Saccharomyces cerevisiae* is

Celsa 822043

reported. UBR1 Encodes a .apprx.225-kDa protein with no significant sequence similarities to other known proteins. Null ubr1 mutants are viable but are unable to degrade the substrates of the N-end rule pathway. These mutants are partially defective in sporulation and grow slightly more slowly than wild-type. The UBR1 protein specifically binds in vitro to proteins bearing amino-terminal residues that are destabilizing according to the N-end rule, but does not bind to otherwise identical proteins bearing stabilizing amino-terminal residues.

IT 136249-46-2

(amino acid sequence of)

=> select hit rn 131

ENTER ANSWER NUMBER OR RANGE (1):1

E1 THROUGH E1 ASSIGNED

=> select hit rn 135 1-2

E2 THROUGH E3 ASSIGNED

=> select hit rn 133 1-10

E4 THROUGH E15 ASSIGNED

=> fil reg

FILE 'REGISTRY' ENTERED AT 09:40:52 ON 25 FEB 93

USE IS SUBJECT TO THE TERMS OF YOUR CUSTOMER AGREEMENT

COPYRIGHT (C) 1993 American Chemical Society (ACS)

STRUCTURE FILE UPDATES: 19 FEB 93 HIGHEST RN 146074-42-2

DICTIONARY FILE UPDATES: 24 FEB 93 HIGHEST RN 146074-42-2

=> s e1

L37 1 111309-64-9/RN

=> s e2-3

1 122983-86-2/RN

1 131571-34-1/RN

L38 2 (122983-86-2/RN OR 131571-34-1/RN)

=> s e4-15

1 130844-14-3/RN

1 136249-46-2/RN

1 137800-99-8/RN

1 139076-24-7/RN

1 139076-28-1/RN

1 142661-94-7/RN

1 142845-03-2/RN

1 143108-32-1/RN

1 143637-87-0/RN

1 143861-32-9/RN

1 143861-35-2/RN

1 144813-79-6/RN

Celsa 822043

L39 12 (130844-14-3/RN OR 136249-46-2/RN OR 137800-99-8/RN OR 139076-24-7/RN OR 139076-28-1/RN OR 142661-94-7/RN OR 142845-03-2/RN OR 143108-32-1/RN OR 143637-87-0/RN OR 143861-32-9/RN OR 143861-35-2/RN OR 144813-79-6/RN)

=> d sqide l37

L37 ANSWER 1 OF 1 COPYRIGHT 1993 ACS
RN 111309-64-9 REGISTRY
CN Antigen 4F2 (human clone pcD-4F2.A heavy chain protein moiety reduced) (9CI) (CA INDEX NAME)
FS PROTEIN SEQUENCE
SSI Antigen 4F2
ORGN human
CLO pcD-4F2.A
SHG reduced
SQL 529

SEQ 1 MSQDTEVDMK EVELNELEPE KQPMNAASGA AMSLAGAEKN GLVKIKVAED
51 EAEAAAAAKF TGLSKEELLK VAGSPGWVRT RWALLLLFWL GWLGMLAGAV
101 VIIVRAPRCR ELPAQKWWHT GALYRIGDLQ AFQGHGAGNL AGLKGRLDYL
151 SSLKVKGVL GPIHKNQKDD VAQTDLLQID PNFGSKEDFD SLLQSAKKKS
201 IRVILDLTNP YRGNSWFST QVDTVATKVK DALEFWLQAG VDGfQVRDIE
251 NLKDASSFLA EWQNITKGFS EDRLLIAGTN SSDLQQILSL LESNKDLLLT
301 SSYLSDSGST GEHTKSLVTQ YLNATGNRWC SWSLSQARLL TSFLPAQLLR
351 LYQLMLFTLP GTPVFSYGDE IGLDAAALPG QPMEAPVMLW DESSFPDIPG
401 AVSANMTVKG QSEDPGSLLS LFRRLSDQRS KERSLLHGDF HAFSAGPGLF
451 SYIRHWDQNE RFLVVLNFGD VGLSAGLQAS DLPASASLPA KADLLLSTQP
501 GREEGSPLLEL ERLKLEPHEG LLLRFPYAA

MF Unspecified

CI MAN

SR CA

LC CA

2 REFERENCES IN FILE CA (1967 TO DATE)

=> d sqide l38 1-2

L38 ANSWER 1 OF 2 COPYRIGHT 1993 ACS
RN 131571-34-1 REGISTRY
CN Protein II (Thermus thermophilus strain HB8 DNA-binding) (9CI) (CA INDEX NAME)
FS PROTEIN SEQUENCE
SSI Protein II
ORGN Thermus thermophilus
SQL 95

SEQ 1 AAKKTVTKAD LVDQVAQATG LKLLDVKAMV DALLAKVEEA LANGSKVQLT
51 GFGTFEVRKR KARTGVKPGT KEKIKIPATQ YPAFKPGKAL KDKVK

MF C458 H781 N127 O129 S

CI MAN

SR CA

Celsa 822043

LC CA
DES 5:ALL,L
1 REFERENCES IN FILE CA (1967 TO DATE)

L38 ANSWER 2 OF 2 COPYRIGHT 1993 ACS

RN 122983-86-2 REGISTRY

CN L-Alaninamide, L-tyrosyl-L-asparaginyL-L-threonyl-L-valyl-L-tyrosyl-L-phenylalanylglycyl-L-phenylalanyl-L-.alpha.-aspartyl-L-lysyl-L-tyrosyl-L-.alpha.-aspartyl-L-isoleucyl-L-threonylglycyl-L-phenylalanyl-L-tyrosyl-L-valyl-L-threonyl-L-isoleucyl-L-.alpha.-aspartyl-L-alanyl-L-.alpha.-aspartyl-L-alanyl-L-alanyl-L-tyrosyl-L-leucyl-L-asparaginyL-L-alanyl-L-threonyl-L-prolyl-L-alanyl- (9CI)
(CA INDEX NAME)

FS PROTEIN SEQUENCE

SSI L-Alaninamide

SBN L-tyrosyl-L-asparaginyL-L-threonyl-L-valyl-L-tyrosyl-L-phenylalanylglycyl-L-phenylalanyl-L-.alpha.-aspartyl-L-lysyl-L-tyrosyl-L-.alpha.-aspartyl-L-isoleucyl-L-threonylglycyl-L-phenylalanyl-L-tyrosyl-L-valyl-L-threonyl-L-isoleucyl-L-.alpha.-aspartyl-L-alanyl-L-.alpha.-aspartyl-L-alanyl-L-alanyl-L-tyrosyl-L-leucyl-L-asparaginyL-L-alanyl-L-threonyl-L-prolyl-L-alanyl-

SQL 33

NTE modified

type	location	description
terminal mod.	Ala-33	- C-terminal amide

SEQ 1 YNTVYFGFDK YDITGFYVTI DADAAYLNAT PAA

MF C173 H241 N37 O52

CI MAN

SR CA

LC CA

DES 5:ALL,L
1 REFERENCES IN FILE CA (1967 TO DATE)

=> d his 140

(FILE 'REGISTRY' ENTERED AT 09:40:52 ON 25 FEB 93)

L40 12 S L12 AND L39

=> d 140 sqide 1-12

L40 ANSWER 1 OF 12 COPYRIGHT 1993 ACS

RN 144813-79-6 REGISTRY

CN Neuraminidase (Actinomyces viscosus strain DSM 43798 reduced) (9CI)
(CA INDEX NAME)

FS PROTEIN SEQUENCE

SSI Neuraminidase

ORGN Actinomyces viscosus

Celsa 822043

SHG reduced
SQL 913

SEQ 1 MTSHPFSRR RLPALLGSLP LAATGLIAAA PPAHAVPTSD GLADVTITQV
51 NAPADGLYSV GDVMTFNITL TNTSGEAHSY APASTNLSGN VSKCRWRNVP
101 AGTTKTDCTG LATHTVTAED LKAGGFTPQI AYEKAVEYA GKALSTPETI
151 KGATSPVKAN SLRVESITPS SSQENYKLGD TVSYTVRVS VSDKTINVAA
201 TESSFDDLGR QCHWGGLKPG KGAVYNCKPL THTITQADVD AGRWTPSITL
251 TATGTDGATL QTLTATGNPI NVVGDHPQAT PAPAPDASTE LPASMSQAQH
301 LAANTATDNY RIPAIPPPPM GTCSSPTTSA RRTTATAAAT TPNPNHIVQR
351 RSTDGGKTWS APTYIHQGTE TGKKVGYS DP SYVVDHQTGT IFNFHVKSYP
401 QGWGGSRRGT DPENRGIIQA EVSTSTDNGW TWTHRTITAD ITKDKPWATAR
== ===
451 FAASGQGIQI QHGPAGRLV QQYTIRTAGG PVQAVSVYSD DHGKTWQAGT
501 PIGTGMDENK VVELSDGSLM LNSRASDGS FRKVAHSTDG GQTWSEPVSD
551 KNLPDSVDNA QIIRAFPNAA PDDPRAKVLL LSHSPNPRPW CRDRGTISMS
601 CDDGASWTTT KVFHEPFVGY TTIAVQSDGS IGLLSEDAHN GADYGGIWYR
651 NFTMNWLGEQ CGQKPAEPSP GRRRRRHPQR HRRRSRPRRP RRALSPRRHR
701 HHPPRPSRAL RPSRAGPGAG AHRSEHGAH TGSCAQSAPE QTDGPTAAPA
751 PETSSAPAAE PTQAPT VAPS VEPTQAPGAQ PSSAPKPGAT GRAPSVVNPK
801 ATGAATEPGT PSSASAPAPS RNAAPT PKPG MEPDEIDRPS DGTMAQPTGA
851 PARRVPRRRR RRRPAAGCLA RDQRAADPGP CGCRGCRRVP AAAGSPFEEL
901 NTRRAGHPAL STD

HITS AT: 419-423

MF Unspecified

CI MAN

SR CA

LC CA

1 REFERENCES IN FILE CA (1967 TO DATE)

L40 ANSWER 2 OF 12 COPYRIGHT 1993 ACS

RN 143861-35-2 REGISTRY

CN Carboxylase, ribulose diphosphate (Bryopsis maxima chloroplast clone
.lambda.-DNA7 gene rbcL large subunit reduced) (9CI) (CA INDEX
NAME)

FS PROTEIN SEQUENCE

SSI Ribulose diphosphate carboxylase

ORGN Bryopsis maxima

CLO .lambda.-DNA7

GEN rbcL

SHG reduced

SQL 474

SEQ 1 APKTETKAGA GFKAGVKDYR LTYYPDYQV KDTDILAAFR MTPQPGVPPE
51 ECGAAVA AES STGTWTTVWT DGLTSLDRYK GRCDLEPVK GEENQYIAYV
101 AYPLDLFE EG SVTNLFTSIV GTVFGFKALR ALRLEDLRIS VAYAKTFQGP
151 PHGIEVERDK LNKYGRPLL G CTIKPKLGLS AKNYGRAVYE CLRGGLDFTK
201 DDENVNSQPF MRWRDRFLFV AEAIYKSQAE TGEIKGHYLN ATAATCEAML
=== ==
251 QRAQCAKELG VPIIMHDYLT GGWTANTSLA HYCRDHGLLL HIHRAMHAVI
301 DRQKNHGMHF RVLAKSLRMS GGDHLHSGTV VGKLEGEREV TLGFVDLMRD

Celsa 822043

351 DFIEKDRARG IYFTQDWVLL PGVMPVASGG IHVWHMPALV EIFGDDACLQ
401 FGGGTLGHPW GNAPGAAANR IACEACTQAR NEGRHLAREG GDVIRAACKW
451 SPELSAACEV LKEIKFEFET IDTL

HITS AT: 238-242

MF Unspecified

CI PMS, MAN

PCT Manual registration

SR CA

LC CA

1 REFERENCES IN FILE CA (1967 TO DATE)

L40 ANSWER 3 OF 12 COPYRIGHT 1993 ACS

RN 143861-32-9 REGISTRY

CN Carboxylase, ribulose diphosphate (Bryopsis maxima chloroplast clone
.lambda.-DNA7 gene rbcL large subunit precursor reduced) (9CI) (CA
INDEX NAME)

FS PROTEIN SEQUENCE

SSI Ribulose diphosphate carboxylase

ORGN Bryopsis maxima

CLO .lambda.-DNA7

GEN rbcL

SHG reduced

SQL 475

SEQ 1 MAPKTETKAG AGFKAGVKDY RLTYYPDPYQ VKDTDILAAF RMTPQPGVPP
51 EECGA AVAAE SSTGTWTTVM TDGLTSLDRY KGRCYDLEPV KGEENQYIAY
101 VAYPLDLFEE GSVTNLFTSI VGTVFGFKAL RALRLEDLRI SVAYAKTFQG
151 PPHGIEVERD KLNKYGRPLL GCTIKPKLGL SAKNYGRAVY ECLRGGLDFT
201 KDDENVNSQP FMRWRDRFLF VAEAIYKSQA ETGEIKGHYL NATAATCEAM
== ==
251 LQRAQCAKEL GVPIIMHDYL TGGWTANTSL AHYCRDHGLL LHIHRAMHAV
301 IDRQKNHGMH FRVLAKSLRM SGGDHLHSGT VVGKLEGERE VTLGFVDLMR
351 DDFIEKDRAR GIYFTQDWVL LPGVMPVASG GIHWHMPAL VEIFGDDACL
401 QFGGTLGHP WGNAPGAAAN RIACEACTQA RNEGRHLARE GGDVIRAACK
451 WPELSAACE VLKEIKFEFE TIDTL

HITS AT: 239-243

MF Unspecified

CI PMS, MAN

PCT Manual registration

SR CA

LC CA

1 REFERENCES IN FILE CA (1967 TO DATE)

L40 ANSWER 4 OF 12 COPYRIGHT 1993 ACS

RN 143637-87-0 REGISTRY

CN Carboxylase, ribulose diphosphate (Bryopsis maxima chloroplast clone
.lambda.MHB7 large subunit precursor reduced) (9CI) (CA INDEX NAME)

FS PROTEIN SEQUENCE

SSI Ribulose diphosphate carboxylase

ORGN Bryopsis maxima

CLO .lambda.MHB7

Celsa 822043

SHG reduced
SQL 475

SEQ 1 MAPKTETKAG AGFKAGVKDY RLTYYPDPYQ VKDTDILAAF RMTPQPGVPP
51 EECGA AVAAE SSTGTWTTVW TDGLTSLDRY KGRCYDLEPV KGEENQYIAY
101 VAYPLDLFEE GSVTNLFTSI VGNVFGFKAL RALRLEDLRI SVAYAKTFQG
151 PPHGIEVERD KLNKYGRPLL GCTIKPKLGL SAKNYGRAVY ECLRGGLDFT
201 KDDENVNSQP FMRWRDRFLF VAEAIYKSQA ETGEIKGHYL NATAATCEAM
== ==
251 LQRAQCAKEL GVPIIMHDYL TGGWTANTSL AHYCRDHGLL LHIHRAMHAV
301 IDRQKNHGMH FRVLAKSLRM SGGDHLHSGT VVGKLEGERE VTLGFVDLMR
351 DDFIEKDRAR GIYFTQDWVL LPGVMPVASG GIHVWHMPAL VEIFGDDACL
401 QFGGGTLGHP WGNAPGAAAN RIACEACVQA RNEGRHLARE GGDVIRAACK
451 WSPELAAACE VWKEIKFEFE TIDTL

HITS AT: 239-243

MF Unspecified

CI MAN

SR CA

LC CA

1 REFERENCES IN FILE CA (1967 TO DATE)

L40 ANSWER 5 OF 12 COPYRIGHT 1993 ACS

RN 143108-32-1 REGISTRY

CN Carboxylase, ribulose diphosphate (sorghum chloroplast clone pSGB510
large subunit reduced) (9CI) (CA INDEX NAME)

FS PROTEIN SEQUENCE

SSI Ribulose diphosphate carboxylase

ORGN sorghum

CLO pSGB510

SHG reduced

SQL 475

SEQ 1 MSPQTETKAS VGFKAGVKDY KLTYYPPEYE TKDTDILAAF RVTPQLGVPP
51 EEAGAAVAAE SSTGTWTTVW TDGLTSLDRV KGRCYHIEPV PGDPDQYICY
101 VAYPLDLFEE GSVTNMFVGI VGNVFGFKAL RALRLEDLRI PPAYLKTFQG
151 PPRGIQIERD KLNKYGRPLL GCTIKPKLGL SAKNYGRACY ECLRGGLDFY
201 KDDENVNSQP FMRWRDRFVF CAEAIYKAQA ETGEIKGHYL NATAGTCEEM
== ==
251 IKRAVFAKEL GVPIVMHDYL TGGFTANTTL SHYCRDNGLL LHIHRAMHAV
301 IDRQKNHGMH FRVLAKALRM SGGDHIHSGT VVGKLEGERE ITLGFVDLLR
351 DFIEKDRSRG IFFTQDWVSM PGVIPVASGG IHVWHMPALT EIIGDDSVLQ
401 FGGGTGLGHPW GNAPGAAANR VALEACVQAR NEGRDLAREG NEIIKAACKW
451 SAELAAACEI WKEIKFDTFK AMDTL

HITS AT: 239-243

MF Unspecified

CI PMS, MAN

PCT Manual registration

SR CA

LC CA

1 REFERENCES IN FILE CA (1967 TO DATE)

Celsa 822043

L40 ANSWER 6 OF 12 COPYRIGHT 1993 ACS
RN 142845-03-2 REGISTRY
CN Cadherin M (mouse C-terminal fragment protein moiety reduced) (9CI)
(CA INDEX NAME)
FS PROTEIN SEQUENCE
SSI Cadherin M
ORGN mouse
SHG reduced
SQL 730

SEQ 1 ENHKRLPYPL VQIKSDKQQL GSVIYSIQGP GVDEEPRNVF SIDKFTGRVY
=

51 LNATLDREKT DRFRLRAFAL DLGGSTLEDP TDLEIVVVDQ NDNRP AFLQD
====

101 VFRGHILEGA IPGTFVTRAE ATDADDPETD NAALRFSILE QGSPEFFSID
151 EHTGEIRTVQ VGLDREVVAV YNLTQVADM SGDGLTATAS AIISIDDIND
201 NAPEFTKDEF FMEAAEAVSG VDVGRLEVED KDLP GSPNWV ARFTILEGDP
251 DGQFKIYTDP KTNEGVLSVV KPLDYESREQ YELRVSVQNE APLQAAAPRA
301 RRGQTRVSVW VQDTNEAPVF PENPLRTSIA EGAPPGTSA TFSARDPDTE
351 QLQRISYSKD YDPEDWLQVD GATGRIQTQR VLSPASPFLK DGWYRAIILA
401 LDNAIPPSTA TGTL SIEILE VNDHAPALAL PPSGSLCSEP DQGPGLLLGA
451 TDEDLPPHGA PFHFQLNPRV PDLGRNWSVS QINVSHARLR LRHQVSEGLH
501 RLSLLLQDSG EPPQQREQTL NVTVCRCGSD GTCLPGAAAL RGGGVGVSLG
551 ALVIVLASTV VLLVLILFAA LRTRFRGHSR GKSL LHGLQE DLRDNILNYD
601 EQGGGEEDQD AYDINQLRHP VEPRATSRSL GRPPLRRDAP FSYVPQPHRV
651 LPTSPSDIAN FISDGLEAAD SDPSVPPYDT ALIYDYE G DG SVAGTLSSIL
701 SSLGDEDQDY DYLRDWGPRF ARLADMYGHQ

HITS AT: 50-54

MF Unspecified

CI MAN

SR CA

LC CA

1 REFERENCES IN FILE CA (1967 TO DATE)

L40 ANSWER 7 OF 12 COPYRIGHT 1993 ACS
RN 142661-94-7 REGISTRY
CN Protein (Drosophila melanogaster clone .lambda.c0.22/.lambda.c0.32
gene small-optic-lobes reduced) (9CI) (CA INDEX NAME)
FS PROTEIN SEQUENCE
SSI Protein
ORGN Drosophila melanogaster
CLO .lambda.c0.22/.lambda.c0.32
GEN small-optic-lobes
SHG reduced
SQL 1597

SEQ 1 MGTISSVLQW SCTKCNTINP TESLKCFNCG TVRKVFPQQQ QQQHRSSSIT
51 ASWTADDALE QEQA EKGQER DKEKGRAAVA RSEYKHVYKS LLRGCLKRPQ
101 RNSQNL PANC VDCEDTRKYI KSSI ELYRHF SNPALNRRWV CHACGTDNSS
151 VTWHCLICDT VSYLAPIYKD AIAADRGQDL AGSLGNRGEL LAADHSHPHH
201 HHHYLHQELE EQHQHQLHSQ HLHKRHLKGR SASGSGSGPG SGSGLRRTQS

Celsa 822043

251	LSTAIDKSAS	GRSCHICYAN	NQSKDIFNLP	QIKPAPQLTG	IPPVAACSNS
301	RFAIANDTFC	RRKQNNNNKN	QNHKVVRESG	AKRKYNFTIT	TLRSAAKDA
351	GHGQMKPLRQ	VVNLNLNLQQ	EPQQKSPANP	QQLHRKTQRE	PAAVSMNPTQ
401	FTIPRNGVFI	AVNEWSEPMA	SSSSVSSSSN	HHHHHSNSN	SNSSGNSNII
451	NNNSSSSSGS	NKLYENECVA	LAQQQLRAAA	AQAAQAAATA	VAIASSPSAK
501	AMAEPAPTAT	MPIYAQVNKQ	HKLKKKQQIA	SESQTNNNTG	SGEIADAVSE
551	SLTAGLGTST	DGSGEASESE	SQVEEHSIYA	KVWKGPBKAT	ESKIMHDPGS
601	SSRLSGAASA	AAGTASAGAI	AAGVGAAAAS	RHDNKTQLGN	GSRSKMWICI
651	KCSYAYNRLW	LQTCEMCEAK	AEQQQQQLHL	QQQQQQQQQH	HHHHHHHLQQ
701	QQAAPRDEP	WTCKKCTLVN	YSTAMACVVC	GGSKLKSIS	IEDMTLRKGE
751	FWTCSHCTLK	NSLHSPVCSA	CKSHRQPQLS	MAMEAVRERP	DGQSYEEQDA
801	AAVGGGGGSA	HQSGANEVKA	PTALNLPLTS	VALPMPMLQL	PTSTAAGLRG
851	SRSPSPRMQL	LPSLQQQRNS	SSSGAIPKRH	STGGSIVPRN	ISIAGLANYN
901	LQQGQGVGSA	SVVSASGAGS	GAGAVGASTS	SKKWQCPACT	YDNCAASVVC
951	DICSSPRGLA	SAVLGEALGR	KSVRVALTPA	DIRQESKLME	NLRQLEETEA
1001	LTKWQNIQY	CRDNSELFVD	DSFPPAPKSL	YYPASGAGE	GNPVVQWRRP
1051	HEINCDGGAY	PPWAVFRTPL	PSDICQGVLG	NCWLLSALAV	LAEREDLVKE
1101	VLVTKEICGQ	GAYQVRLCKD	GKWTTVLVDD	LLPCDKRGHL	VYSQAKRKQL
1151	WVPLIEKAVA	KIHGCYEALV	SGRAIEGLAT	LTGAPCESIP	LQASSLPMP
1201	EDELDKDLIW	AQLLSSRCVR	FLMGASCGGG	NMKVDEEEYQ	QKGLRPRHAY
1251	SVLDVKDIQG	HRLLKLRNPW	GHYSWRGDWS	DDSSLWTDDL	RDALMPHGAS
1301	EGVFWISFED	VLNYFDCIDI	CKVRSGWNEV	RLQGTLLQPLC	SISCVLLTVL
		=====			
1351	EPTEAEFTLF	QEGQRNSEKS	QRSQDLDCVV	IFRTRSPAAP	EIGRLVEHSK
1401	RQVRGFVGCH	KMLERDIYLL	VCLAFNHWH	GIEDPHQYPQ	CILAIHSSKR
1451	LLVEQISPSP	HLLADAIISL	TLTKGQRHEG	REGMTAYYLT	KGWAGLVVMV
1501	ENRHENKWIH	VKCDQCQESYN	VVSTRGELKT	VDSVPPLQRQ	VIIIVLTQLEG
1551	SGGFSIAHRL	THRLANSRGL	HDWGPPGATH	CPPIENVHGL	HAPRLIT

HITS AT: 1311-1315

MF Unspecified

CI MAN

SR CA

LC CA

1 REFERENCES IN FILE CA (1967 TO DATE)

L40 ANSWER 8 OF 12 COPYRIGHT 1993 ACS

RN 139076-28-1 REGISTRY

CN Protein NS 3 (tick-borne encephalitis virus strain 205 clone I-47/III-30/I-30 reduced) (9CI) (CA INDEX NAME)

FS PROTEIN SEQUENCE

SSI Protein NS 3

ORGN tick-borne encephalitis virus

CLO I-47/III-30/I-30

SHG reduced

SQL 621

SEQ	1	SDLVFSGQGG	RERGDRPFEV	KDGVYRIFSP	GLLWGQRQVG	VGYGSKGV LH
	51	TMWHVTRGAA	LSIDDAVAGP	YWADV KEDVV	CYGGAWSLEE	KWKGETVQVH
	101	AFPPGRAHEV	HQCQPGELLL	DTGRRIGAVP	IDLAKGTSGS	PILNSQGVVV
	151	GLYGNGLKTN	ETYVSSIAQG	EAEKSRPNLP	PAVSGTGWTA	KGQITVLDMH
	201	PGSGKTHRVL	PELIRQCTDR	RLRTLVLAPT	RVVLKEMERA	LNGKRVRFHS

Celsa 822043

251 PAVGDQQVGG AIVDVMCHAT YVNRRLLPQG RQNWEVAIMD EAHWTDPHSI
 301 AARGHLYTLA KENKCALVLM TATPPGKSEP FPESNGAISS EEKQIPDGEW
 351 RDGFDWITEY EGRTAWFVPS IVKGGIIART LRQKGKSVIC LNSKTFEKDY
 401 SRVRDEKPDF VVTTDISEMG ANLDVSRVID GRTNIKPEEV DGRVELTGTR
 451 RVTTASAAQR RGRVGRQEGR TDEYIYSGQC DDDDSGLVQW KEAQILLDNI
 501 TTLRGPVATF YGPEQDKMPE VAGHFRLTEE KRKHFRHLLT HCDFTPWLAW

=

551 HVAANVSSVT SRNWTWEGPE ENTVDEANGD LVTFRSPNGA ERTLRPVWRD

=====

601 ARMFREGRDI REFVAYASGR R

HITS AT: 550-555

MF Unspecified

CI MAN

SR CA

LC CA

1 REFERENCES IN FILE CA (1967 TO DATE)

L40 ANSWER 9 OF 12 COPYRIGHT 1993 ACS

RN 139076-24-7 REGISTRY

CN Protein, poly- (tick-borne encephalitis virus strain 205 clone
 II-71/I-62/I-3/I-47/III-30/I-30/III-15/II-4/III-45/II-6/VI-49
 reduced) (9CI) (CA INDEX NAME)

FS PROTEIN SEQUENCE

SSI Polyprotein

ORGN tick-borne encephalitis virus

CLO II-71/I-62/I-3/I-47/III-30/I-30/III-15/II-4/III-45/II-6/VI-49

SHG reduced

SQL 3414

SEQ 1 MAGKAILKGK GGGPPRRVSK ETAKKTRQSR VQMPNGLVLM RMMGILWHAV
 51 AGTARSPVLK SFWNSVPLKQ ATAALRKIKK AVSTLMVGLQ RRGKRRSAVD
 101 WTGWLLVVVL IGVTLAATVR KERDGTTVIR AEGKDAATQV RVENGTCVIL
 151 ATDMGSWCDD SLTYECVTID QGEEPVDVDC FCRNVDGVYL EYGRCGKQEG
 201 SRTRRSVLIP SHAQGDLTGR GHKWLEGDSL RTHLTRVEGW VWKNKILTLA
 251 VIAVWLTVE SVVTRIAVVV VLLCLAPVYA SRCTHLENRD FVTGTQGTTR
 301 VTLVLELGGC VTITAEGKPS MDVWLDSIYQ ENPAKTREYC LHAKLSDTKV
 351 AARCPTMGPA TLAEHHQSGT VCKRDQSDRG WGNHCGLFGK GSIVTCVKAS
 401 CEAKKKATGH VYDANKIVYT VKVEPHTGDY VAANETHSGR KTASFTVSSE
 451 KTILTMGDYG DVSLLCRVAS GVDLAQTVIL ELDKTSEHLP TAWQVHRDWF
 501 NDLALPWRHE GAQNWNNNAER LVEFGAPHAV KMDVYNLGDQ TGVLLKSLAG
 551 VPVAHIDGTK YHLKSGHVTC EVGLEKLKMK GLTYTMCDKT KFTWKRTPTD
 601 SGHDTVMEV AFSGTKPCRI PVRAVAHGSP DVNVAMLITP NPTIENNGGG
 651 FIEMQLPPGD NIIYVGELSH QWFQKGSSIG RVFQKTRKGI ERLTVIGEHA
 701 WDFGSTGGFL TSVGKALHTV LGGAFNSLFG GVGFLPKILM GVALAWLGLN
 751 MRNPTMSMGF LLAGGLVLAM TLGVGADVGC AVDTERMELR CGEGLVVWRE
 801 VSEWYDNYAY YPETPGALAS AIKETFEEGT CGIVPQNRLE MAMWRSSATE
 851 LNLALAEGBA NLTVVVDKLD PTDYRGGIPG LLRKGKDIKV SWKSWGHSMI
 901 WSVPEAPRRF MVGTEGSSEC PLERRKTGVF TVAIEFGVGLR TKVFLDFRQE
 951 PTHECDTGVM GAAVKNGMAV HTDQSLWMKS VRNDTGTYIV ELLVTDLRNC
 1001 SWPASHTIDN AEVVDSELFL PASLAGPRSW YNRIPGYSEQ VKGPWKYSPI
 1051 RVTREECPGT RVTINADCDK RGASVRSTTE SGKVIPEWCC RTCTLPPVTF

Celsa 822043

1101	RTGTDCWYAM	EIRPVHDQGG	LVRSMVVADN	GELLSEGGIP	GIVALFVVLE
1151	YVIRRRPATG	TTAMWGGIVV	LALLVTGLVK	IESLVRYVVA	VGITFHLELG
1201	PEIVALTLLQ	AVFELRVGLL	SAFALRSNLT	VREMTIYFL	LLVLELGLPS
1251	EGLGALWKWG	DALAMGALIF	RACTAEKGTG	VGLLLMALMT	QODLAIHYG
1301	LMLFLGVASC	YSIWKLIRGH	REQKGLTWIV	PLAGLLGGEG	SGVRLAFWE
1351	LAIHGRRRSF	SEPLTVVGVM	LTLASGMMRH	TSQEALCALA	VASFLLMLV
1401	LGTRKMQLVA	EWSGCWEVHP	ELMNEGGEVS	LRVRQDSMGN	FHLTELEKEE
1451	RVMAFWLLAG	LAASAFHWSG	ILGVMGLWTL	SEMLRTARRS	DLVFSGQGR
1501	ERGDRPFVK	DGVYRIFSPG	LLWGQRQGV	GYGSKGVLHT	MWHVTRGAAL
1551	SIDDAVAGPY	WADVKEDVVC	YGGAWSLEEK	WKGETVQVHA	FPPGRAHEVH
1601	QCQPGELLD	TGRRIGAVPI	DLAKGTSOSP	ILNSQGVVVG	LYGNGLKTNE
1651	TYVSSIAQGE	AEKSRPNLPP	AVSGTGWTAK	GQITVLDMHP	GSGKTHRVLP
1701	ELIRQCTDRR	LRTLVLAPTR	VVLKEMERAL	NGKRVRFHSP	AVGDQQVGGA
1751	IVDVMCHATY	VNRRLLPQGR	QNWEVAIMDE	AHWTDPHSIA	ARGHLYTLAK
1801	ENKCALVLMT	ATPPGKSEPF	PESNGAISSE	EKQIPDGEWR	DGFDWITEYE
1851	GRTAWFVPSI	VKGGIIARTL	RQKGKSVICL	NSKTFEKDYS	RVRDEKPDFV
1901	VTTDISEMGA	NLDVSRVIDG	RTNIKPEEVD	GRVELTGTRR	VTTASAAQRR
1951	GRVGRQEGRT	DEYIYSGQCD	DDDSGLVQWK	EAQILLDNIT	TLRGPVATFY
2001	GPEQDKMPEV	AGHFRLTEEK	RKHFRHLLTH	CDFTPWLAWH	VAANVSSVTS

== ==

2051	RNWTWEGPEE	NTVDEANGDL	VTFRSPNGAE	RTLRPVWRDA	RMFREGRDIR
2101	EFVAYASGRR	SFGDVLGMS	GVPELLRHRC	VSAMDVFYTL	MHEEPGRAM
2151	KMAERDAPEA	FLTVAEMMVL	GLATLGVVWC	FVVRTSISRM	MLGTLVLLAS
2201	LALLWAGGVS	YGNMAGVALI	FYTLLTVLQP	EAGKQRSSDD	NKLAYFLLTL
2251	CSLAGLVAA	EMGFLEKTKA	DLSTVLWSEH	EELRSWEWT	NIDIQPARSW
2301	GTYYLVVSLF	TPYMIHQLOT	KIQQLVNSAV	VTGAQAMRDL	GGGAPFFGIA
2351	GHVMALGVVS	LVGATPTSIV	VGVGLAAFHL	AIVVSGLEAE	LTQRAHKVFF
2401	SAMVRNPMVD	GDVINPFGE	EAKPALYERK	MSLVLAIVLC	LMSVVMNRTV
2451	PSITEASAVG	LAAAGQLLRP	EVDTLWTMPV	ACGLSGVVRG	SLWGFLPLGH
2501	RLWLRASGSR	RGGSEGDTLG	DLWKRKLNGC	TKEEFFAYRR	TGILETERDK
2551	ARELLRRGET	NMGLAVSRGT	AKLAWLEERG	YATLKGEVVD	LGCGRGWSY
2601	YAASRPVMS	VKAYTIGGKG	HETPRMVTSL	GWNLKIFRAG	MDVFSMQPHR
2651	ADTIMCDIGE	SNPDVVEGE	RTRKVILLME	QWKNRNPTAT	CVFKVLAPYR
2701	PEVIEALHRF	QLQWGGGLVR	TPFSRNSTHE	MYYSTAVTGN	IVNSVNIQSR
2751	KLLARFGDQR	GPTRVPELDL	GVGTRCVVLA	EDKVKEKDVQ	ERISALREQY
2801	GETWHMDREH	PYRTWQYWGS	YRTAPTGSAA	SLINGVVKLL	SWPWNAREDV
2851	VRMAMTDTTA	FGQQRVFKEK	VDTKAQEPQP	GTKVIMRAVN	DWILERLARK
2901	SKPRMCSREE	FIKVKSNAA	LGAWSDEQNR	WSSAKEAVED	PAFWQLVDEE
2951	RERHLAGRCA	HCVYNMMGKR	EKKLGFEFGA	KGSRAIWMW	LGSRLFLEFEA
3001	LGFLNEDHWA	SRGFSGSGVE	GISLNYLGWY	LKELSTLEGG	LFYADDTAGW
3051	DTKVTNADLE	DEEQLLRYME	GEHRQLAATI	MQKAYHAKVV	KVARPSRDGG
3101	CIMDVITRRD	QRGSGQVVTY	ALNTLTNIKV	QLIRMMEGEG	VIEASDAHNP
3151	RLLRVERWLR	DHGEERLGRM	LVSGDDCVVR	PVDDRFRGRAL	YFLNDMAKTR
3201	KDIGEWEHSV	GFSNWEEVPF	CSHHFHELVM	KDGRALIVPC	RDQDELVGRA
3251	RVSPGCGWSI	RETACLSKAY	GQMWLLSYFH	RRDLRTLGLA	ICSAVPVDWV
3301	PTGRTTWSIH	ASGAWMTTED	MLDVWNRVWI	LDNPFMHSKE	KIVEWRDPY
3351	LPKSHDMLCS	SLVGRKERAE	WAKNIWGAVE	KVRRMIGQEK	FKDYLSCMDR
3401	HDLHWELKLE	SSII			

HITS AT: 2039-2044

MF Unspecified

CI MAN

Celsa 822043

SR CA
LC CA

1 REFERENCES IN FILE CA (1967 TO DATE)

L40 ANSWER 10 OF 12 COPYRIGHT 1993 ACS
RN 137800-99-8 REGISTRY
CN Protein IBMP (figwort mosaic virus clone DxS reduced) (9CI) (CA INDEX NAME)
FS PROTEIN SEQUENCE
SSI Protein IBMP
ORGN figwort mosaic virus
CLO DxS
SHG reduced
SQL 512

SEQ 1 MEELKALRLK EKILEIELNS VKMQIHAYEE SLKATTVNSV QEGEILQTES
51 IPECPAQGKE TPNPVKADSL LKTILGNERQ NPLEGKSSKL VNLTPKSDKD
101 KVKSSPVANG SGKDSTKPLN PVALGKSKMT ILGQKQADEE EFKPDYLRRAA
151 SNGQSWFAVY KGPNKEFFTE WEIVADICK RQKSKRFRSK EQAEVSISLY
=====

201 NKDIQDPVNF LRPVKLVKEE RAAQPLKFKA IAAEQTIQFD EFRQIWEKSR
251 LSDLEDGVQE KFYTNDASAK STYTFVENAE PYLVHTAFRA GLAKVIYPSP
301 NLQELKWFPE GIVKAIKNFR KKVLNAKDAA IFIKIFSSIP DWVQSTRYEP
351 YHFIQIGIAK TKKELPSSKV CKEEFSVQSL NKVRVQSLQT ISQKLQEINE
401 ESSIKVNYCS STCIMVSKFQ KKTSTEDLKL VGIFESNLVN IEQLACGDQT
451 KKEWCRIVR TYQKHLCLYC KDKADSSSTS GEQNNVEKSC PDSPLTNAYD
501 ERSDDHKRIP SI

HITS AT: 192-196

MF Unspecified

CI MAN

SR CA

LC CA

1 REFERENCES IN FILE CA (1967 TO DATE)

L40 ANSWER 11 OF 12 COPYRIGHT 1993 ACS
RN 136249-46-2 REGISTRY
CN Synthetase, ubiquitin-protein (Saccharomyces cerevisiae clone pUBR1 reduced) (9CI) (CA INDEX NAME)
FS PROTEIN SEQUENCE
SSI Ubiquitin-protein synthetase
ORGN Saccharomyces cerevisiae
CLO pUBR1
SHG reduced
SQL 1950

SEQ 1 MSVADDDLGS LQGHIRRTL R SIHNLPHYFRY TRGPTERADM SRALKEFIYR
51 YLYFVISNSG ENLPTLFNAH PKQKLSNP EL TVFPDSLEDA VDIKITSQQ
101 TIPFYKIDES RIGDVHKHTG RNCGRKFKIG EPLYRCHECG CDDTCVLCIH
151 CFNPKDHVNH HVCTDICTEF TSGICDCGDE EAWNSPLHCK AEEQENDISE
201 DPATNADIKE EDVWNDSVNI ALVELVLA EV FDYFIDVFNQ NIEPLPTIQK
251 DITIKLREMT QQGKMYERAQ FLNDLKYEND YMFDGTTTAK TSPSNSPEAS

Celsa 822043

301 PSLAKIDPEN YTVIIYNDEY HNYSQATTAL RQGVDPNVHI DLLTSRIDGE
351 GRAMLKCSQD LSSVLGGFFA VQTNGLSATL TSWSEYLHQE TCKYIILWIT
401 HCLNIPNSSF QTTFRNMMGK TLCSEYLNAT ECRDMTPVVE KYFSNKFDDN

=====

451 DPYRYIDLSI LADGNQIPLG HHKILPESST HSLSPINDV ETPTSRTYSN
501 TRLQHILYFD NRYWKRLRKD IQNVIIPTLA SSNLYKPIFC QQVVEIFNHI
551 TRSVAYMDRE POLTAIRECV VQLFTCPTNA KNIFENQSFL DIVWSIIDIF
601 KEFCKVEGGV LIWQRVQKSN LTKSYSISFK QGLYTVETLL SKVHDPNIPL
651 RPKEIISLLT LCKLFNGAWK IKRKEGEHVL HEDQNFISYL EYTTISIYSII
701 QTAEKVSEKS KDSIDSKLFL NAIRIISFL GNRSLTYKLI YDSHEVIKFS
751 VSHERVAFMN PLQTMLSFLI EKVSCLKDAYE ALEDSCDFLK ISDFSLSRVV
801 LCSQIDVGFW VRNGMSVLHQ ASYYKNNPEL GSYSRDIHLN QLAILWERDD
851 IPRIIYNILD RWELLDWFTG EVDYQHTVYE DKISFIIQQF IAFIYQILTE
901 RQYFKTFSSL KDRRMDQIKN SIIYNLYMKP LSYSKLLRSV PDYLTEDTTE
951 FDEALEEVSF FVEPKGLADN GVFKLKASLY AKVDPLKLLN LENEFESEAT
1001 IIKSHLAKDK DEIAKVVLIP QVSIKQLDKD ALNLGAFTRN TVFAKVVKYL
1051 LQVCLDMEDS TFLNELLHLV HGIFRDELI NGKDSIPEAY LSKPICNLLL
1101 SIANAKSDVF SESIVRKADY LLEKMIMKKP NLFESLIAS FGNQYVNDYK
1151 DKKLRQGVNL QETEKERKRR LAKKHQARLL AKFNNQQTKF MKEHESEFDE
1201 QDNDVDMVGE KVEYSEDFTC ALCQDSSSTD FFVIPAYHDH SPIFRPGNIF
1251 NPNEFMPMWD GFYNDDEKQA YIDDDVLEAL KENGSCGSRK VVSCNHHIH
1301 HNCFKRYVQK KRFSSNAFIC PLCQTFSNCT LPLCQTSKAN TGLSLDMFLE
1351 SELSLDTLR LFKPFTEENY RTINSIFSLM ISQCQGFDA VRKRANFSHK
1401 DVSLILSVHW ANTISMLEIA SRLEKPYSIS FFRSREQKYK TLKNILVCIM
1451 LFTFVIGKPS MEFEPYPQQP DTWVNQNQLF QYIVRSALFS PVSLRQTVTE
1501 ALTTFSRQFL RDFLQGLSDA EQVTKLYAKA SKIGDVLKVS EQMLFALRTI
1551 SDVRMEGLDS ESIIYDLAYT FLLKSLPTI RRCLVFIKVL HELVKDSENE
1601 TLVINGHEVE EELEFEDTAE FVNKALKMIT EKESLVDLLT TQESIVSHPY
1651 LENIPYEYCG IIKLIDLSKY LNTYVTQSKE IKLREERSQH MKNADNRDLF
1701 KICLTCGVKV HLRADRHMT KHLNKNCFKP FGAFMPNSS EVCLHLTQPP
1751 SNIFISAPYL NSHGEVGRNA MRRGDLTTLN LKRYEHLNRL WINNEIPGYI
1801 SRVMGDEFVR TILSNGFLFA FNREPRPRI PPTDEDDDM EGEDGFFTE
1851 GNDEMVDVDE TGQAANLFGV GAEGIAGGGV RDFFQFFENF RNTLQPQNG
1901 DDDAPQNPP ILQFLGPQFD GATIIRNTNP RNLEDDSD NDDSDEREIW

HITS AT: 426-430

MF Unspecified

CI PMS, MAN

PCT Manual registration

SR CA

LC CA

1 REFERENCES IN FILE CA (1967 TO DATE)

L40 ANSWER 12 OF 12 COPYRIGHT 1993 ACS

RN 130844-14-3 REGISTRY

CN Protein (potato virus X strain Xc 165-kilodalton reduced) (9CI) (CA
INDEX NAME)

FS PROTEIN SEQUENCE

SSI Protein

ORGN potato virus X

SHG reduced

SQL 1456

Celsa 822043

```

SEQ      1 MAKVREYVQS FTDSTTKTLI QDEAYRNIRP IMEKHKLSNP YAQTIEAAND
      51 LEGFGIATNP YSIELHTHAA AKTIENKLE VLGSLLPQEP VTFMFLKPRK
     101 LNFMRNPRI KDIFHNVAIE PRDVARYPKE TIIHKLAIE TDTAYISDTL
     151 HFLDPSYIVE TFQNCPLQT LYATLVLPSE AAFKMESTHP NIYSLKYFGD
     201 GFQYIPGNHG GGAYHHEFTD LQWLKVGKIK WRDPKDGLLG HLNYPHEQVD
     251 THTVTVQLQE SFAANHLYCI RRGNMMPTEV RTFGQPDYV LPPQIILPKV
     301 HNCKKPILKK TMMQLFLYVR TVKVAKNCID FAKVRQLIKS SDRDKFSAVE
     351 LVYLVSYMEF LAALQATTCF SDTSLGGLLT KTLAPVRAWI QEKKMQLCGL
     401 EDYVKLVKAV DWRPVDFSFK VETWDFRFTQ LGMWKAFQPS ELSDVEEMNN
     451 FFDDGDLDC FTRMPAYAVN AEEDLAGMRS NNQEETSTAP REPEGEKKEY
     501 INPAETFLDK LTRKHNRER SRAAKKAKRL AEIQDSMNRD RTEEGSHKTP
     551 NMGEAPSNAD LPGVNEVEAG TTFTPLKALP QKWEDASSTD SSTIDPTEII
     601 PGEEDDKAAT QKVVGGLPWK HWLPQLNAV GFKALEIQDR NGTMIMPITE
     651 MVFGLDKEEF PEGTPEALAR ELKAMNRSPT TIPLDLLRAR DYGSVDKNKR
     701 IGAITKTQAA SWGEYLTGKI ESLPERKVAA CVIHGAGGSG KSHAIQKALR
     751 EIGKGS DITV VLPTNELRLD WSKKVPNTEP YMFKTYEKAL IGGTGSIVIF
     801 DDYSKLPPGY IEALVSFSTK IKLIILT GDS RQSVYHETSD DASIRHLGPA
     851 TEVFAKYCRY YLNATHRNKK DLANMLGVYS ERTGTTEISM SSEFLEGVPT
          =====
     901 LVPSDEKRRL YMGTRNDTF TYAGCQGLTK PKVQIVLDHN TQVCSANVMY
     951 SALSRA TDRI NFINTSANSS AFWEKLDSTP YLKTFLSVVR EHALKEYEPA
    1001 EA EPIKEPEP QTHMCVENEE SVLEEYKEEL LEKFDREIHS DAHGHSNCVQ
    1051 TEDTTIQLFS HQEAKDETL L WATIDARLKT SNQESNFREF LSKRDIGDVL
    1101 FLNYQKAMGL PKEPIPF SQE VWEACAHEVQ SKYLSKSKCN LINGTVRQSP
    1151 DFDENKIMVF LKSQWVTKVE KLGLPKIKPG QTIAAFYQQT VMLFGTMARY
    1201 MRWFRQAFQP KEVFINCETT PEDMSAWALS NWNFTRPSLA NDYTAFDQSQ
    1251 DGAMLQFEVL KAKHHCIPEE IIQAYIDIKT HAQIFLG YLS IMRLTGEGPT
    1301 FDANTECNIA FTHYKFDIPA GTAQVYAGDD SALDCVPEVK QSFHRLDKL
    1351 LLKSKPVITQ QKKGSWPEFC GWLITPKGVM KDPIKLHVSL KLAEAKGELR
    1401 KCQDSYEIDL SYAYDHKDSL NDLFDEKQCQ AHTLT CRTLI KSGRGTVSLP
    1451 RLKNFL

```

HITS AT: 861-865

MF Unspecified

CI MAN

SR CA

LC CA

2 REFERENCES IN FILE CA (1967 TO DATE)